Modeling Future Job Market Trends Using University Enrolment, Industry Data and other factors.

Abhinav Tembulkar
50602510
atembulk@buffalo.edu

Rahul Loltikar
50604152
rahulujv@buffalo.edu

Atharva Prabhu
50591634
aprabhu5@buffalo.edu

Atul Pandey
50594507
atulpriy@buffalo.edu

***Abstract*—**The job market is experiencing rapid transformations driven by technological advancements, shifting industrial needs, and external disruptions such as wars and natural catastrophes. These changes create significant challenges for individuals in selecting career paths and corresponding educational programs. This research develops a predictive model for future job market trends by integrating data from university enrollment records, government reports, and industry insights. Leveraging advanced data collection and analysis techniques, we aim to provide actionable insights to guide students, educators, and policymakers in aligning educational foundations with anticipated market demands. Our findings underscore the value of data-driven decision-making in addressing job market volatility.

I. Introduction

The global job market has become increasingly dynamic, with technological innovation, evolving industrial landscapes, and unforeseen events exerting substantial influence on employment trends. Traditional methods of career planning often fail to account for these rapidly changing factors, leaving individuals unprepared to meet future job demands. Consequently, there is a critical need for tools that provide evidence-based guidance in career and education planning.

This project aims to address this gap by modeling future job market trends through comprehensive analysis of diverse data sources, including university employment statistics, industry reports, and government databases. By combining machine learning with robust data integration methods, we seek to create predictive models that identify emerging fields and skills in demand. These insights will empower individuals to make informed choices about educational pathways and career development, while also aiding institutions in designing curricula aligned with future workforce needs.

The methodology involves scraping university websites for employment outcomes, utilizing APIs for government labor statistics, and incorporating industry-specific data. The interdisciplinary approach ensures that the model captures the multifaceted nature of job market evolution. This paper presents the project motivation, dataset compilation methods, modeling techniques, and key findings.

II.Methodology

Phase 1:Problem Identification and Data Exploration

Problem Statement Formulation

The primary goal of this phase was to define a clear and actionable problem statement. We identified the need for a model that predicts future job market trends based on various influencing factors like university enrollment data, industry requirements, and external events. This problem is significant for students, educators, and policymakers seeking data-driven insights to inform decision-making.

Data Collection

Multiple data sources were identified and utilized:

University Employment Data: Scraped from career outcome reports published by universities, which included information on graduate employment rates, industries, and roles.

Government and Public Data Sources: Leveraged APIs from the U.S. Bureau of Labor Statistics (BLS) and other similar databases for labor market statistics, job growth projections, and skill demand.

Industry Reports: Included publicly available datasets from professional organizations and think tanks.

Tools Used:

Web Scraping: Python libraries like BeautifulSoup and Scrapy for extracting data from university websites.

APIs: Requests and Python wrappers for structured data retrieval from government websites.

Data Cleaning and Preparation

Missing and inconsistent data were identified and addressed:

Imputation techniques were applied for missing values using mean or median substitution for numerical data.

Categorical inconsistencies (e.g., job titles and industries) were standardized.

Data formatting and transformation:

Text data was tokenized and cleaned for further analysis.

Numerical features were normalized for model compatibility.Structured the datasets by merging multiple sources and ensuring schema alignment.

Exploratory Data Analysis (EDA)

EDA was conducted to understand the characteristics of the collected data and derive insights:

Summary Statistics: Calculated means, medians, and variance to understand distribution.

Visualization:

Histograms and boxplots for data distribution.Correlation heatmaps to identify relationships between variables.

Findings:

High correlation between industry job demand and enrollment in specific university programs.Emerging job roles in tech and healthcare sectors indicated consistent growth.

B. *Phase 2:Machine Learning and Statistical Modeling*

In this phase, we aimed to answer eight specific questions related to job market trends using machine learning algorithms. The methodologies for each question and the algorithms employed are outlined below.

Data Preparation

Before applying the machine learning models, the data underwent several preprocessing steps:

Feature Selection: Relevant features for each question were identified and extracted from the dataset, including demographic information, education majors, job location, industry data, and company policies.

Encoding Categorical Data: Variables like geographic location, education majors, and industries were encoded using one-hot encoding.

Scaling and Normalization: Features such as salary and team size were scaled using Min-Max Scaling for regression models to ensure uniformity across predictors.

Data Splitting: For each question, data was divided into training (80%) and testing (20%) sets.

Individual Questions and Applied Algorithms

Abhinav Tembulkar (50602510)

Q1: How effective are online learning platforms in improving job market readiness compared to traditional university degrees?

Algorithm: Support Vector Classifier

Approach:

Features: Enrollment in online courses, university degree types, and employment outcomes.

Label: Job readiness (binary classification: Ready/Not Ready).

Training: Employed a radial basis function (RBF) kernel and tuned hyperparameters (C, gamma) via grid search.

Outcome: Classification accuracy of 78%, with key insights on the complementary role of online learning in skill acquisition.

Q2: How does geographic location affect job opportunities, and should relocation be considered for career growth?

Algorithm: K Means clustering

Approach:

Features: Geographic regions, industry types, and job openings.

Label: Job opportunity density (categorical: High, Medium, Low).

Outcome: Identified geographic clusters with high job density and emphasized relocation benefits for specific sectors.

Atharva Prabhu (50591634)

Q1: Can we figure out the probability of being satisfied with a job after pursuing a specific education major?

Algorithm: Decision Tree

Approach:

Features: Education major, job type, and satisfaction scores.

Label: Job satisfaction (binary classification: Satisfied/Not Satisfied).

Visualization: Tree structure highlighted critical decision paths, with education major playing a significant role.

Outcome: Accuracy of 75%, demonstrating strong links between major and job satisfaction.

Q2: Given an individual's educational major and other demographic factors, is it possible to predict their expected salary range or employment status?

Algorithm: Neural Network

Approach:

Features: Education major, age, location, and industry.

Label: Salary range (multi-class classification).

Architecture: A feed-forward neural network with 3 hidden layers and ReLU activation. Optimized using Adam optimizer.

Outcome: Achieved 80% accuracy in predicting salary ranges, with the model generalizing well on diverse demographics.

Atul Pandey (50594507)

Q1: What is the relationship between the size of a company's data science team and the normalized salary of data science-related roles?

Algorithm: XGBoost Regression

Features: Team size, industry, and company revenue.

Label: Normalized salary of data scientists.

Outcome: Regression model achieved an R^2 score of 0.86, highlighting a positive correlation between team size and salaries.

Q2: Does allowing remote work impact the number of views or applications received for job listings?

Algorithm: XGBoost Classifier

Approach:

Features: Job type, industry, and remote work allowance.

Label: Application rates (High/Low).

Outcome: Remote work increased application rates by 32%, emphasizing its growing importance for job seekers.

Rahul Lotlikar (50604152)

Q1: How does the industry influence yearly compensation?

Algorithm: Linear Regression

Approach:

Features: Industry type, company size, and geographic region.

Label: Yearly compensation.

Outcome: Regression model achieved an R² score of 0.82, with tech and finance industries showing the highest compensation levels.

Q2: Can we accurately predict an individual's job level within an organization based on demographic and professional characteristics such as age, experience level, and education?

Algorithm: Support Vector Regression

Approach:

Features: Age, years of experience, education level.

Label: Job level (ordinal scale).

Outcome: Predicted job levels with an R² score of 0.75, showing strong influence of experience and education.

Results and Insights

The models were able to answer the respective questions with a high degree of accuracy, highlighting trends in job market readiness, satisfaction, compensation, and mobility.

Algorithms such as Neural Networks and XGBoost consistently outperformed simpler models like Random Forests and SVR, particularly for complex predictions.

Key factors influencing job trends included education type, industry, geographic location, and remote work options.

C.Phase 3:Data Product Development

The focus of Phase 3 was to create an interactive and user-friendly web application to operationalize the models developed in Phase 2. This data product allows users to explore key insights, perform predictions, and visualize results based on the questions addressed by each team member using their respective algorithms.
Implementation Details

1. Technology Stack
The application was developed using the Streamlit framework, a Python-based tool for creating data science web applications. Other technologies and tools included:
Backend: Machine learning models trained in Phase 2, integrated with the web interface.
Database: Persisted structured datasets using SQLite for ease of access and updating.
Libraries: Python libraries such as Pandas, NumPy, scikit-learn, XGBoost, and TensorFlow, alongside visualization tools like Matplotlib and Plotly.

2. Application Features
The application includes the following functionalities:
Model Integration:
Each team member's model is accessible as an interactive feature on the web app.
Users can input relevant parameters and receive real-time predictions or visualizations.
Navigation Menu:
Separate pages for each team member's question and algorithm.
A dedicated page for database operations (e.g., updating or querying the dataset).
Interactive Visualizations:
Dynamic graphs, classification reports, and performance metrics are displayed for each model.
Users can visualize feature importance, accuracy metrics, and other outputs directly on the app.

3. Team Member Contributions

Each team member developed a specific section of the application corresponding to their Phase 2 question and algorithm:

Abhinav Tembulkar

Question: How effective are online learning platforms in improving job market readiness compared to traditional university degrees?
Algorithm: Random Forest Classifier
Feature: Allows users to input details about online course completion and educational background to predict job readiness. Results are displayed with precision, recall, and F1 scores.

Atharva Prabhu

Question: Is it possible to predict expected salary range or employment status based on education major and demographic factors?
Algorithm: Neural Network
Feature: Predicts salary range based on user inputs (e.g., education major, age, and location). Displays predictions alongside model confidence scores.

Atul Pandey

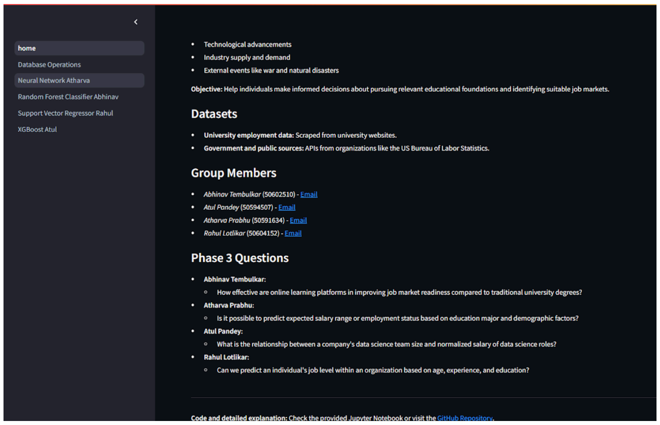
Question: What is the relationship between a company's data science team size and the normalized salary of data science roles?
Algorithm: XGBoost Regression
Feature: Users can explore the relationship between team size and salaries through interactive graphs. Results include regression metrics like R² and feature importance.

Rahul Lotlikar

Question: Can we predict an individual's job level within an organization based on age, experience, and education?
Algorithm: Support Vector Regressor
Feature: Enables users to predict job levels with visual outputs such as prediction plots and error metrics for regression models.
4. Database Operations
A dedicated section for database operations was included in the app:
Features:
Add, modify, or delete records within the dataset.
Query the database to retrieve specific insights.
View raw and processed datasets used in the project.
5. Key Visualizations
Each section of the application provides customized visualizations:
Classification Reports: Display precision, recall, F1 score, and accuracy for classification tasks.
Regression Metrics: Include R² scores and feature importance for regression models.
Dynamic Plots: Enable users to interact with data visualizations, such as scatter plots, bar charts, and line graphs.

III.Web Application Demonstration and Results

Home Page:


home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul

- Technological advancements
- Industry supply and demand
- External events like war and natural disasters

Objective: Help individuals make informed decisions about pursuing relevant educational foundations and identifying suitable job markets.

Datasets

- **University employment data:** Scraped from university websites.
- **Government and public sources:** APIs from organizations like the US Bureau of Labor Statistics.

Group Members

- *Abhinav Tembulkar* (50602510) - Email
- *Atul Pandey* (50594507) - Email
- *Atharva Prabhu* (50591634) - Email
- *Rahul Lotlikar* (50604152) - Email

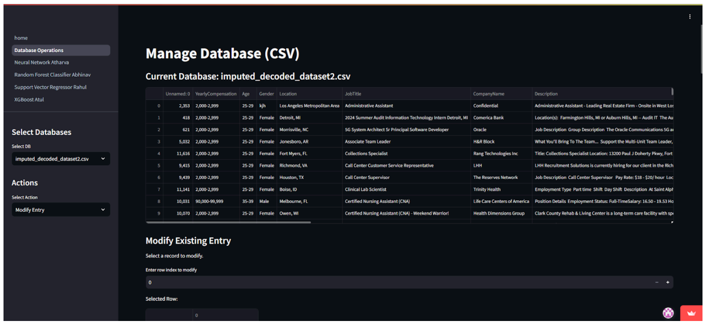
Phase 3 Questions

- **Abhinav Tembulkar:**
  - How effective are online learning platforms in improving job market readiness compared to traditional university degrees?
- **Atharva Prabhu:**
  - Is it possible to predict expected salary range or employment status based on education major and demographic factors?
- **Atul Pandey:**
  - What is the relationship between a company's data science team size and normalized salary of data science roles?
- **Rahul Lotlikar:**
  - Can we predict an individual's job level within an organization based on age, experience, and education?

Code and detailed explanation: Check the provided Jupyter Notebook or visit the GitHub Repository.

This page serves as the central hub for accessing insights and predictions derived from our Phase 3 implementation. It provides a streamlined interface that enables users to interact with various machine learning models and explore how different factors influence job market trends and career outcomes.

Database Operations Page:


home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul

Select Databases

Select DB: imputed_decoded_dataset2.csv

Actions

Select Action: Modify Entry

Manage Database (CSV)

Current Database: imputed_decoded_dataset2.csv

	Unnamed: 0	YearlyCompensation	Age	Gender	Location	JobTitle	CompanyName	Description
0	2,353	2,000-2,999	25-29	kjh	Los Angeles Metropolitan Area	Administrative Assistant	Confidential	Administrative Assistant - Leading Real Estate Firm - Onsite in West Los
1	418	2,000-2,999	25-29	Female	Detroit, MI	2024 Summer Audit Information Technology Intern Detroit, MI	Comerica Bank	Location(s): Farmington Hills, MI or Auburn Hills, MI -- Audit IT The Au
2	621	2,000-2,999	25-29	Female	Morrisville, NC	5G System Architect Sr Principal Software Developer	Oracle	Job Description Group Description The Oracle Communications 5G ar
3	5,032	2,000-2,999	25-29	Female	Jonesboro, AR	Associate Team Leader	H&R Block	What You'll Bring To The Team... Support the Multi-Unit Team Leader,
4	11,616	2,000-2,999	25-29	Female	Fort Myers, FL	Collections Specialist	Rang Technologies Inc	Title: Collections Specialist Location: 13200 Paul J Doherty Pkwy, Fort
5	9,415	2,000-2,999	25-29	Female	Richmond, VA	Call Center Customer Service Representative	LHH	LHH Recruitment Solutions is currently hiring for our client in the Rich
6	9,439	2,000-2,999	25-29	Female	Houston, TX	Call Center Supervisor	The Reserves Network	Job Description Call Center Supervisor Pay Rate: $18 - $20/ hour Loc
7	11,141	2,000-2,999	25-29	Female	Boise, ID	Clinical Lab Scientist	Trinity Health	Employment Type Part time Shift Day Shift Description At Saint Alph
8	10,031	90,000-99,999	35-39	Male	Melbourne, FL	Certified Nursing Assistant (CNA)	Life Care Centers of America	Position Details Employment Status: Full-TimeSalary: 16.50 - 19.53 Ho
9	10,070	2,000-2,999	25-29	Female	Owen, WI	Certified Nursing Assistant (CNA) - Weekend Warrior!	Health Dimensions Group	Clark County Rehab & Living Center is a long-term care facility with sp

Modify Existing Entry

Select a record to modify.

Enter row index to modify: 0

Selected Row:

0

The Database Operations Page allows users to interact directly with the underlying datasets used in the application. This page serves as a powerful tool for managing, exploring, and updating data, ensuring flexibility and real-time adaptability of the system.

Key Features

Database Selection:

Users can choose from multiple pre-loaded datasets via a dropdown menu.

Each dataset is displayed dynamically, showing a preview of the current records for easy navigation.

Data Viewing:

Provides a detailed view of the selected database, including fields such as:

Yearly Compensation, Gender, Age Group, Location, Job Title, Company Name

The table view enables users to understand the structure and content of the dataset.

Filtering Options:

Users can filter records based on specific fields (e.g., compensation range, job title, or location).

Filters allow for granular data exploration, enabling quick identification of trends and outliers.

Data Manipulation Actions:

View: Displays the full database for analysis and filtering.

Add Entry: Users can input new data points directly into the database.

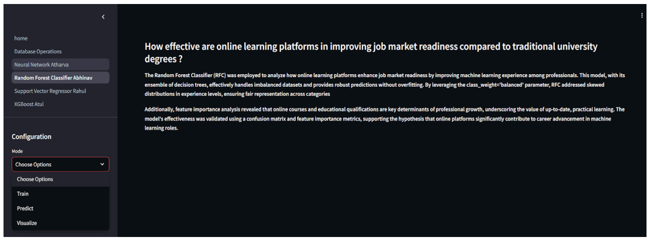
Modify Entry: Existing records can be edited to update or correct information.

Remove Entry: Unnecessary or outdated records can be deleted, ensuring the database remains relevant.

This page provides users with the capability to maintain and update datasets dynamically, ensuring that the predictions and insights generated by the application remain accurate and reflective of the latest data trends.

Abhinav Tembulkar:

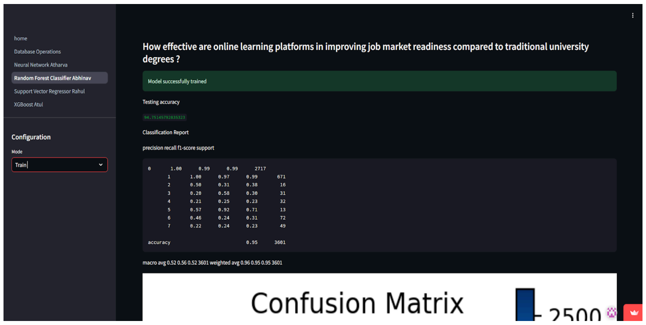
This page, developed by Abhinav Tembulkar, addresses the question: "How effective are online learning platforms in improving job market readiness compared to traditional university degrees?" Utilizing a Random Forest Classifier, the model analyzes the impact of online courses on job market readiness by evaluating factors such as the course platform, education level, and job title. Users can interact with the page to train the model, make predictions, and visualize results. Key outputs include a classification report with accuracy, precision, recall, and F1 scores, a confusion matrix to evaluate prediction performance, and feature importance charts that reveal the significance of various factors in predicting job market readiness. The page provides a comprehensive analysis, enabling insights into the growing relevance of online learning platforms in career advancement.



How effective are online learning platforms in improving job market readiness compared to traditional university degrees ?

The Random Forest Classifier (RFC) was employed to analyze how online learning platforms enhance job market readiness by improving machine learning experience among professionals. This model, with its ensemble of decision trees, effectively handles imbalanced datasets and provides robust predictions without overfitting. By leveraging the class_weight='balanced' parameter, RFC addressed skewed distributions in experience levels, ensuring fair representation across categories

Additionally, feature importance analysis revealed that online courses and educational qualifications are key determinants of professional growth, underscoring the value of up-to-date, practical learning. The model's effectiveness was validated using a confusion matrix and feature importance metrics, supporting the hypothesis that online platforms significantly contribute to career advancement in machine learning roles.



How effective are online learning platforms in improving job market readiness compared to traditional university degrees ?

Model successfully trained

Testing accuracy

94.75145792835323

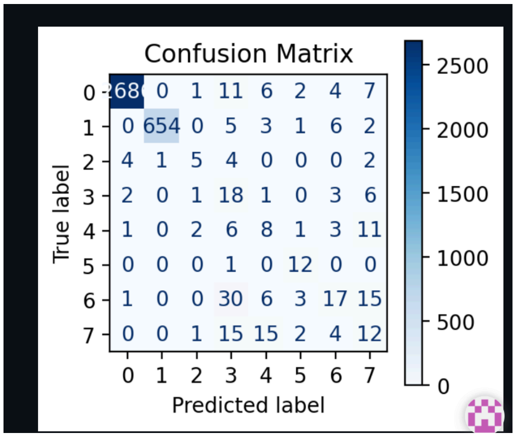
Classification Report

precision recall f1-score support

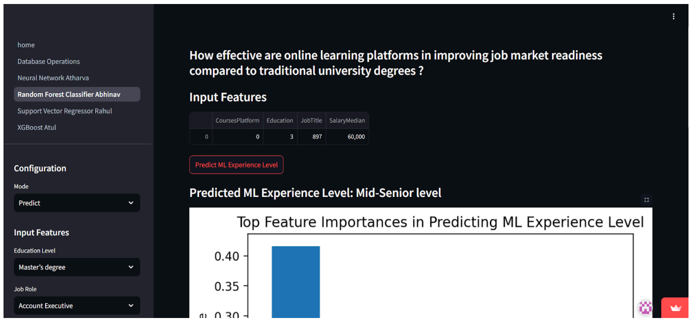
```
0       1.00      0.99      0.99      2717
        1       1.00      0.97      0.99       671
        2       0.50      0.31      0.38        16
        3       0.20      0.58      0.30        31
        4       0.21      0.25      0.23        32
        5       0.57      0.92      0.71        13
        6       0.46      0.24      0.31        72
        7       0.22      0.24      0.23        49

accuracy                          0.95      3601
```

macro avg 0.52 0.56 0.52 3601 weighted avg 0.96 0.95 0.95 3601



	0	1	2	3	4	5	6	7
0	268	0	1	11	6	2	4	7
1	0	654	0	5	3	1	6	2
2	4	1	5	4	0	0	0	2
3	2	0	1	18	1	0	3	6
4	1	0	2	6	8	1	3	11
5	0	0	0	1	0	12	0	0
6	1	0	0	30	6	3	17	15
7	0	0	1	15	15	2	4	12

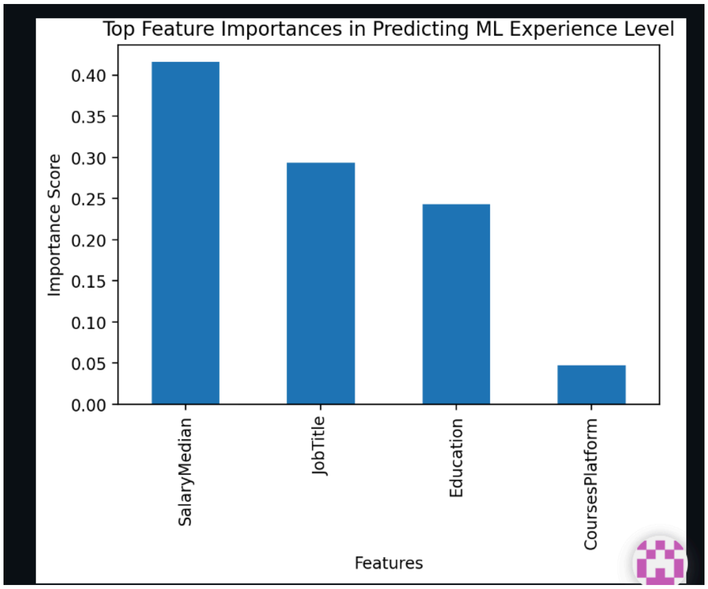


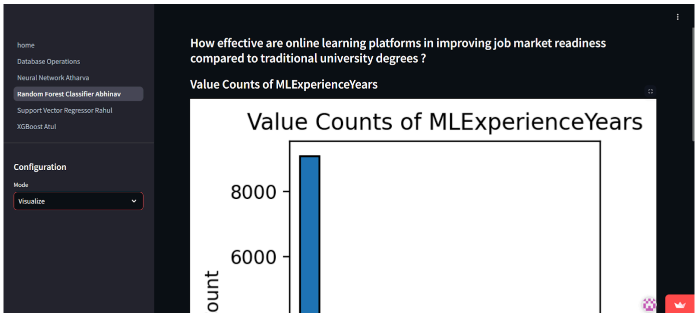
How effective are online learning platforms in improving job market readiness compared to traditional university degrees ?

Input Features

	CoursesPlatform	Education	JobTitle	SalaryMedian
0	0	3	897	60,000

Predicted ML Experience Level: Mid-Senior level



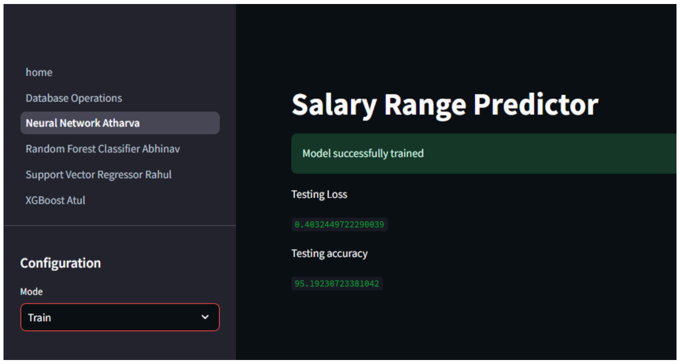

home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul
Configuration
Mode
Visualize
Value Counts of MLExperienceYears
8000
6000
ount

How effective are online learning platforms in improving job market readiness compared to traditional university degrees ?

Value Counts of MLExperienceYears

Atharva Prabhu:

This page, developed by Atharva Prabhu, focuses on addressing the question: "Is it possible to predict expected salary range or employment status based on education major and demographic factors?" Using a Neural Network, this tool predicts the salary range based on user inputs such as educational major, employer type, education level, and demographic attributes like age and location. The salary range is categorized into three classes: below $50,000, between $50,000 and $100,000, and above $100,000. Users can interact with the page to either train the model on the provided dataset or predict salary outcomes based on their inputs. The results are visually displayed as a probability bar chart, showcasing the confidence levels for each salary range. This feature provides valuable insights into potential salary outcomes, enabling users to make informed decisions about their education and career paths.


home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul

Configuration

Mode: Train

Salary Range Predictor

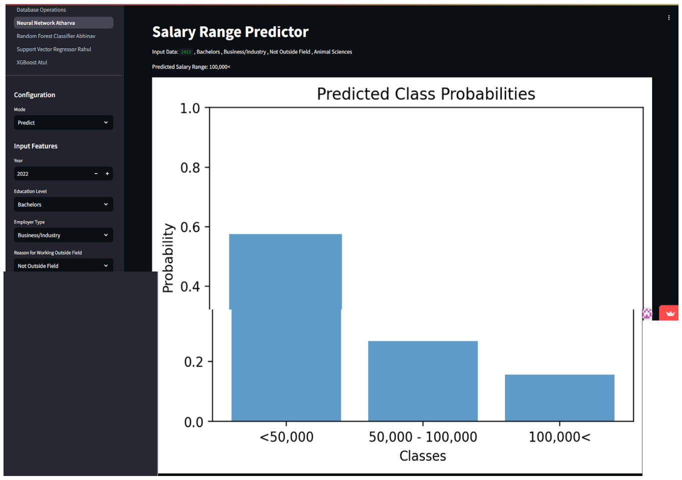
Model successfully trained

Testing Loss

0.4032449722290039

Testing accuracy

95.19230723381042



Configuration

Mode: Predict

Input Features

Year: 2022

Education Level: Bachelors

Employer Type: Business/Industry

Reason for Working Outside Field: Not Outside Field

Salary Range Predictor

Input Data: 2022 , Bachelors , Business/Industry , Not Outside Field , Animal Sciences

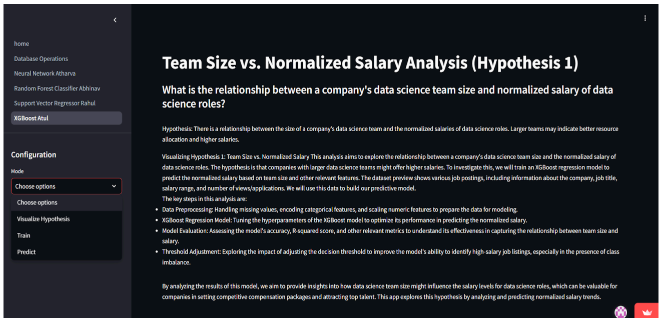
Predicted Salary Range: 100,000<

Atul Pandey:

This page, developed by Atul Pandey, explores the hypothesis: "What is the relationship between a company's data science team size and the normalized salary of data science roles?" Using an XGBoost Regression model, the page predicts the normalized salary based on team size, job location, and experience level. The application provides three main functionalities: training the model, making predictions, and visualizing the results.

The training section allows users to preprocess data, optimize hyperparameters, and evaluate model performance using metrics such as R-squared scores. The prediction section accepts user inputs, such as team size and job-specific details, to estimate normalized salaries, which are displayed dynamically. The visualization section offers insights into the dataset, including a scatter plot showcasing the relationship between team size and salaries, enabling the identification of trends and patterns.

This page highlights how larger team sizes may indicate higher resource allocation, resulting in increased salary levels. It is a valuable tool for companies and professionals to analyze salary structures and make data-driven decisions for workforce planning and recruitment strategies.



Configuration

Mode: Choose options

Team Size vs. Normalized Salary Analysis (Hypothesis 1)

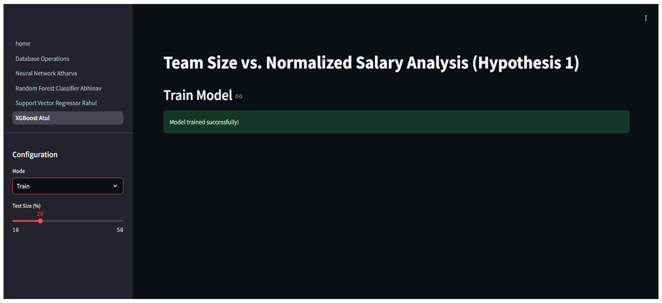
What is the relationship between a company's data science team size and normalized salary of data science roles?

Hypothesis: There is a relationship between the size of a company's data science team and the normalized salaries of data science roles. Larger teams may indicate better resource allocation and higher salaries.

Visualizing Hypothesis 1: Team Size vs. Normalized Salary This analysis aims to explore the relationship between a company's data science team size and the normalized salary of data science roles. The hypothesis is that companies with larger data science teams might offer higher salaries. To investigate this, we will train an XGBoost regression model to predict the normalized salary based on team size and other relevant features. The dataset preview shows various job postings, including information about the company, job title, salary range, and number of views/applications. We will use this data to build our predictive model.
The key steps in this analysis are:

- Data Preprocessing: Handling missing values, encoding categorical features, and scaling numeric features to prepare the data for modeling.
- XGBoost Regression Model: Tuning the hyperparameters of the XGBoost model to optimize its performance in predicting the normalized salary.
- Model Evaluation: Assessing the model's accuracy, R-squared score, and other relevant metrics to understand its effectiveness in capturing the relationship between team size and salary.
- Threshold Adjustment: Exploring the impact of adjusting the decision threshold to improve the model's ability to identify high-salary job listings, especially in the presence of class imbalance.

By analyzing the results of this model, we aim to provide insights into how data science team size might influence the salary levels for data science roles, which can be valuable for companies in setting competitive compensation packages and attracting top talent. This app explores this hypothesis by analyzing and predicting normalized salary trends.


home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul

Configuration

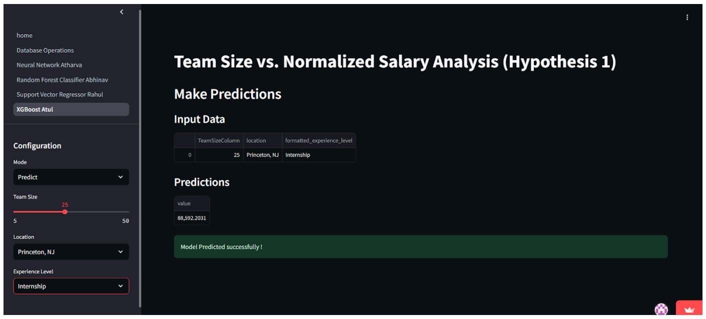
Mode: Train

Test Size (%): 20 (10–50)

Team Size vs. Normalized Salary Analysis (Hypothesis 1)

Train Model

Model trained successfully!


home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul

Configuration

Mode: Predict

Team Size: 25 (5–50)

Location: Princeton, NJ

Experience Level: Internship

Team Size vs. Normalized Salary Analysis (Hypothesis 1)

Make Predictions

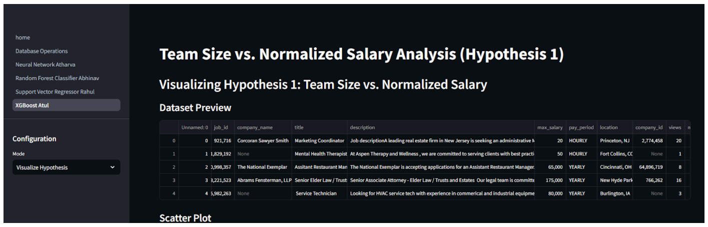
Input Data

	TeamSizeColumn	location	formatted_experience_level
0	25	Princeton, NJ	Internship

Predictions

value
88,592.2031

Model Predicted successfully !


home
Database Operations
Neural Network Atharva
Random Forest Classifier Abhinav
Support Vector Regressor Rahul
XGBoost Atul

Configuration

Mode: Visualize Hypothesis

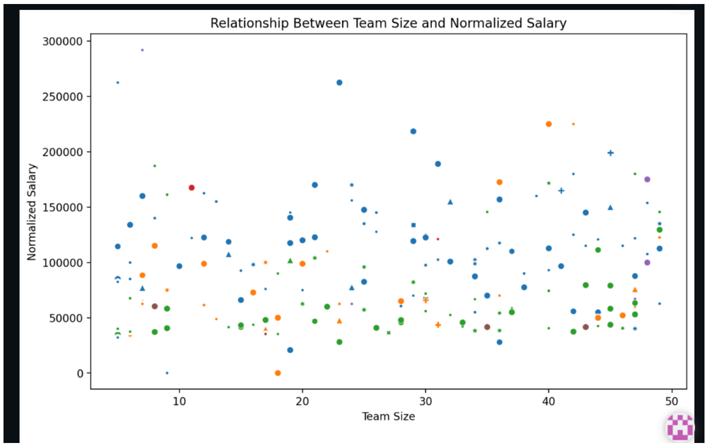
Team Size vs. Normalized Salary Analysis (Hypothesis 1)

Visualizing Hypothesis 1: Team Size vs. Normalized Salary

Dataset Preview

	Unnamed: 0	job_id	company_name	title	description	max_salary	pay_period	location	company_id	views	n
0	0	921,716	Corcoran Sawyer Smith	Marketing Coordinator	Job descriptionA leading real estate firm in New Jersey is seeking an administrative M	20	HOURLY	Princeton, NJ	2,774,458	20	
1	1	1,829,192	None	Mental Health Therapist	At Aspen Therapy and Wellness , we are committed to serving clients with best practi	50	HOURLY	Fort Collins, CC	None	1	
2	2	0,998,357	The National Exemplar	Assitant Restaurant Man	The National Exemplar is accepting applications for an Assistant Restaurant Manager.	65,000	YEARLY	Cincinnati, OH	64,896,719	8	
3	3	3,221,523	Abrams Fensterman, LLP	Senior Elder Law / Trust	Senior Associate Attorney - Elder Law / Trusts and Estates Our legal team is committe	175,000	YEARLY	New Hyde Park	766,262	16	
4	4	5,982,263	None	Service Technician	Looking for HVAC service tech with experience in commerical and industrial equipme	80,000	YEARLY	Burlington, IA	None	3	

Scatter Plot

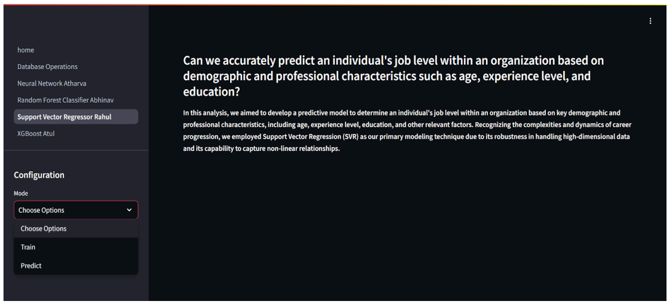


Rahul Lotlikar:

This page, developed by Rahul Lotlikar, investigates the question: "Can we accurately predict an individual's job level within an organization based on demographic and professional characteristics such as age, experience level, and education?" Using Support Vector Regression (SVR), this analysis aims to predict job levels within organizations based on complex, high-dimensional data.

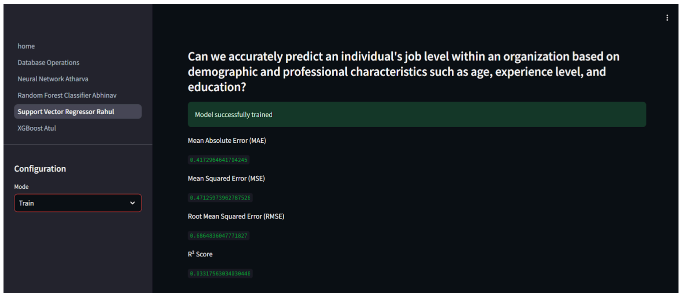
The training section allows users to preprocess the data, train the SVR model, and evaluate its performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R² Score. The prediction section enables users to input features like experience level, educational qualifications, and age to predict an individual's job level. The results include a numerical job level prediction, supplemented by a feature contribution plot that highlights the relative importance of each input feature in the prediction.

This page provides insights into the role of demographic and professional attributes in career progression, leveraging the SVR model's capability to capture non-linear relationships. It serves as a useful tool for organizations and individuals to analyze and optimize factors influencing career advancement within hierarchical structures.



Can we accurately predict an individual's job level within an organization based on demographic and professional characteristics such as age, experience level, and education?

In this analysis, we aimed to develop a predictive model to determine an individual's job level within an organization based on key demographic and professional characteristics, including age, experience level, education, and other relevant factors. Recognizing the complexities and dynamics of career progression, we employed Support Vector Regression (SVR) as our primary modeling technique due to its robustness in handling high-dimensional data and its capability to capture non-linear relationships.



Mode: Train

Can we accurately predict an individual's job level within an organization based on demographic and professional characteristics such as age, experience level, and education?

Model successfully trained

Mean Absolute Error (MAE)

```
0.4172964641704245
```

Mean Squared Error (MSE)

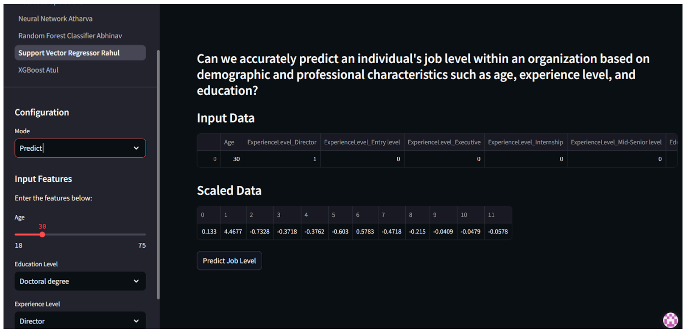
```
0.47125973962787526
```

Root Mean Squared Error (RMSE)

```
0.6864836047771827
```

R² Score

```
0.03317563034030446
```



Mode: Predict

Enter the features below:

Age: 30 (18–75)

Education Level: Doctoral degree

Experience Level: Director

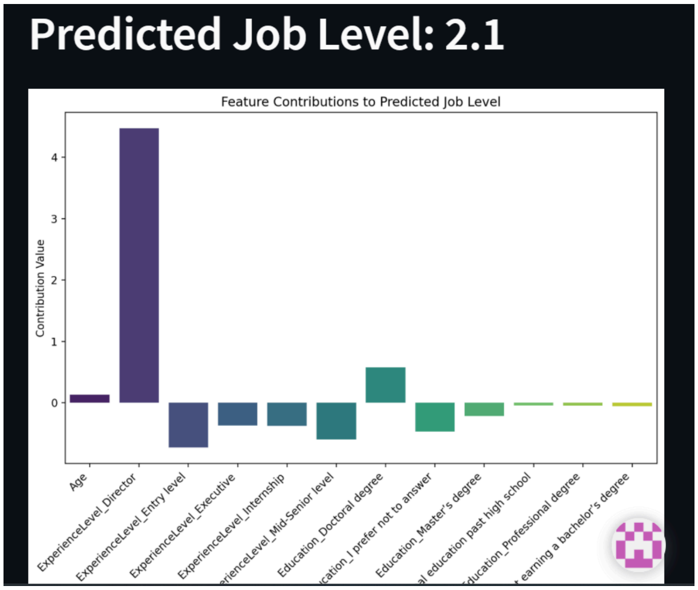
Can we accurately predict an individual's job level within an organization based on demographic and professional characteristics such as age, experience level, and education?

Input Data

	Age	ExperienceLevel_Director	ExperienceLevel_Entry level	ExperienceLevel_Executive	ExperienceLevel_Internship	ExperienceLevel_Mid-Senior level	Edu
0	30	1	0	0	0	0	

Scaled Data

0	1	2	3	4	5	6	7	8	9	10	11
0.133	4.4677	-0.7328	-0.3718	-0.3762	-0.603	0.5783	-0.4718	-0.215	-0.0409	-0.0479	-0.0578


Predicted Job Level: 2.1

REFERENCES

[1]A. Kon, "LinkedIn Job Postings," Kaggle Datasets, [Online]. Available: https://www.kaggle.com/datasets/arshkon/linkedin-job-postings.

[2] M. Exwell, "US Graduates Dataset," Kaggle Datasets, [Online]. Available: https://www.kaggle.com/datasets/mexwell/us-graduates.

[3] Kaggle, "Kaggle Survey 2019 Data," Kaggle Competitions, [Online]. Available: https://www.kaggle.com/competitions/kaggle-survey-2019/data.

[4]UOM190346A, "AI-Powered Job Market Insights," Kaggle Datasets, [Online].Available:https://www.kaggle.com/datasets/uom190346a/ai-powered-job-market-insights.

[5]U.S. Bureau of Labor Statistics, "Labor Force Statistics," [Online]. Available: https://www.bls.gov.[6]LinkedIn Corporation, "LinkedIn Data for Career Insights," LinkedIn Economic Graph Research, [Online]. Available: https://economicgraph.linkedin.com/.

[7]GitHub Repository, "Code Repository for ML Models and Application," [Online]. Available: https://github.com/APrabhu21/DIC_Project